

## FUTURE PROSPECTS IN COMMUNITY DETECTION

MINI SINGH AHUJA<sup>1</sup> & JATINDER SINGH<sup>2</sup>

<sup>1</sup>Research Scholar, Punjab Technical University Jalandhar, Punjab, India

<sup>2</sup>Professor, KC Group of Institutes Nawashahr, Punjab, India

### ABSTRACT

In recent years Internet usage has increased. People of all age group use Internet and this has led to the invention of a new research field called complex networks. Complex networks such as social networks, biological networks, technological networks etc have become the interest of many researchers because of their wide range of applications. These complex networks have many properties like scale free networks, transitivity etc but the most important property is the presence of community structure in these networks. In this paper we have studied the various properties of complex networks giving more emphasis to community structure. In this paper we tried to collect all the information related to community detection and also provided future work directions in this area.

**KEYWORDS:** Communities, Clusters, Graph Theory, Complex Networks, Community Detection

### INTRODUCTION

Now day's real systems have grown in size tremendously. They contain million of actors and have different relationships. Complex networks are the powerful modelling tools which represent most real world systems. Complex network paradigm is one of the modelling tools which have spread through several application fields such as sociology, communication, computer science, biology, and physics and so on during last decades. Complex networks can be represented in the form of large graphs which have large no of nodes and different types of relationships with non trivial properties. These nodes can be anything: a person, an organization, a computer or a biological cell. Nodes can have different size or attributes which represent a property of real system objects. These graphs can be directed, undirected or weighted. A complex network has its roots in graph theory. Few examples of complex networks are Internet maps (IP, Routers [30], web graphs (hyperlinks between pages) [15], data exchange (emails) [31], social networks (facebook, twitter, scientist collaboration networks) biological networks (protein interaction, epidemic networks) etc. Complex networks have non trivial properties so they cannot be explained by uniform random, regular or complete models [29]. This has resulted in definition of set of statistics which have become fundamental properties of complex networks. These properties are now being used by many researchers for studying various phenomena's like spreading of information [19], protocol performance etc. But a major challenge in the study of complex networks is how to collect data for analysis. We cannot directly collect data from these real world complex networks to study them. So researches have to make an assumption that initially data is not fit to find the real properties but as the size of the data grows the properties become more and more stable. The research is going on this side of complex network too [31]. They are trying to find the impact of the measured procedures on the obtained data to study the induced bias [32].

## PROPERTIES OF COMPLEX NETWORKS

Complex networks are very popular these days because they are flexible and have a very natural structure. Complex networks have many properties which have been investigated in the past. These are:

- **Small World Effect**

A network is said to have small world property if the average path length between the nodes is less. This concept has originated from the famous experiment which was carried out by Milgram in 1967. In his experiment he realised that the letters that were passed from person to person were able to reach a destination target individually in small six steps. Small world effect is also called “the six degree of separation” principle. The most popular model of random networks with small world characteristics was developed by Watts and Strongatz and is called Watts and Strongatz (WS) small world model [2]. Many real world networks such as road maps, metabolic networks, social influence networks show small world property. In many real world networks for a fixed average degree, average path length increases non linearly with network size but then also they follow the small world property.

- **Scale Free Network**

Networks which have power law degree distribution are called scale free networks. Many real world networks such as WWW, biological networks, social networks, citation networks are scale free networks. Scale-free networks were introduced by Barabasi and Albert. Scale [6] free networks are highly vulnerable to a coordinated attack against their hubs. A scale-free network can be constructed by adding nodes to an existing network and links to existing nodes with preferential attachment so that the probability of linking to a given node  $i$  is proportional to the number of existing links  $k_i$  that node has, i.e.,

$$P(\text{linking to node } i) \sim \frac{k_i}{\sum_j k_j}.$$

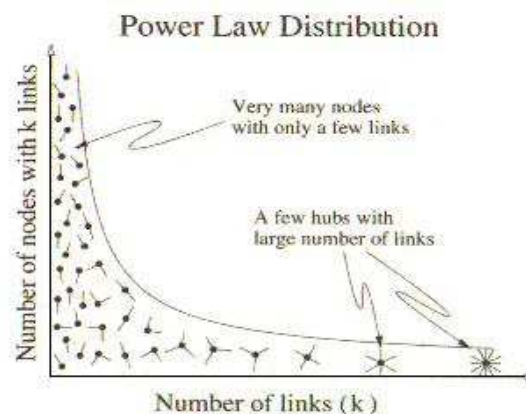


Figure 1

- **Assortativity**

The scale-free metric is related to the assortativity coefficient  $r$ , which is a measure of the likelihood for nodes to connect to other nodes with similar degrees. The assortativity is calculated by calculating the Pearson correlation coefficient between the degrees of all pairs of nodes connected by an edge. The value of ' $r$ ' lies between -1 and 1. If we a

high assortativity coefficient, it means that nodes tend to connect to nodes of similar degree, but if we have a negative coefficient, it means that nodes are likely to connect to nodes with very different degree from their own.

$$p_k \propto k^{-\alpha}.$$

- **Community Structure**

Many real world networks like social networks and biological networks have a modular structure [1]. These networks are formed by set of vertices called communities. These structures have more connections within the same community than between different communities. The first model to generate such networks with the community structure property was proposed by Girvan and Newman

- **Clustering Coefficient**

Clustering coefficient is defined as the average fraction of pair of neighbours of a node that are also neighbours of each other. Suppose if a node  $i$ 's neighbours have  $n$  directed links between them, then the clustering coefficient of  $i$  is defined as

$$c(i) = \frac{n}{d_i(d_i - 1)}$$

The clustering coefficient of a graph is the average clustering coefficient of all its nodes, and it is represented as  $C(G)$ ,

$$C(G) = \frac{\sum_{v \in V} c(v)}{|V|}$$

The clustering coefficient of a graph lies between 0 and 1. Higher values indicate that there is higher degree of “cliquishness” between the nodes. If a graph has clustering coefficient value 0 it means it has no “triangles” of connected nodes, but if a graph has clustering coefficient of 1 is a perfect clique.

- **Degree Distribution**

Degree distribution  $P(k)$  gives the probability of a node to have a degree  $k$  in the network. The number of nodes with degree  $k$  can be denoted as  $n_k$ . So degree distribution is

$$P(k) = \frac{n_k}{N}$$

Many real world networks follow the power law degree distribution

- **Density**

The total number of nodes usually tells us about the size of the network. Density is defined as the level of linkage between the nodes. It is calculated as the ratio of the number of existing links to the number of possible links. In a complete network, all the nodes are connected to all other nodes. So the density is 1.

$$d = \frac{m}{n(n-1)}$$

Where  $m$  is no of existing nodes and  $n$  no of possible nodes

- **Node Centrality**

There are different measures to determine the importance of nodes in the network. Networks can be directed, undirected or weighted. Various type of centrality measures are:

**Betweenness Centrality:** It is defined as the number of number of shortest path running from the given node.

$$C_B(n_i) = \sum_{j < k} g_{jk}(n_i) / g_{jk}$$

Where  $g_{jk}$  = the number of geodesics connecting  $jk$ , and  $g_{jk}(n_i)$  = the number that actor  $i$  is on.

**Closeness Centrality:** A node is considered important if it is relatively close to other nodes. It is the inverse of the sum of the distances between this node and all other nodes.

$$C_c(n_i) = \left[ \sum_{j=1}^g d(n_i, n_j) \right]^{-1}$$

Where  $d(n_i, n_j)$  is the distance from node  $i$  to node  $j$  and  $i$  the given node.

**Degree centrality:** It is defined as the no of edges of a node.

$$C = \frac{\text{Vertex Degree}}{n-1}$$

**Eigen Vector Centrality:** Eigenvector centrality measures the influence of a [node](#) in a [network](#). It assigns relative scores to all nodes in the network.

## COMMUNITIES

Large real world networks are generally characterized by heterogeneous structures which have some particular properties. The heterogeneous distribution of the links has led to community structure [1, 4, 18]. A community can be described as a collection of vertices within graph which are densely connected among themselves but are loosely connected to the rest of the graph (Newman, and Girvan, 2004). Communities can also be called as clusters or modules which share common properties. These communities have many features. They can have hierarchal or overlapping structure inside them. Moreover these communities can be dynamic which change with time or can be multirelational (multiple relations). Many real networks such as social networks, biological networks exhibit community structure. This property of complex networks can be used in various applications such as to study the spread of disease in social networks [10]. Web clients who have similar interests and are geographically near to each other can be clustered to improve the performance of the service providers on the World Wide Web. Each cluster can be served by a dedicated mirror server. Community Structure property reduces very large graph in to smaller ones. Now days, community detection has become a popular field of research. Community detection algorithms are the common and fundamental tools which help to uncover the principles present in networks. Community detection algorithms focus only towards the network structure. While detecting the communities, two possible sources of information are expected: network structure, and the attributes and features of the nodes. Many algorithms for community detection have come up till now.

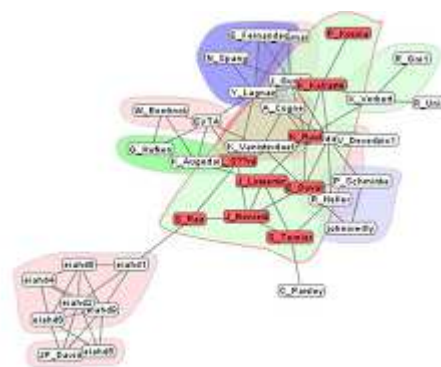


Figure 2: Communities

## COMMUNITY DETECTION VS CLUSTER ANALYSIS

Community detection is one of the widely discussed problems in network science and it has a wide range of applications [1]. These applications include detecting friend circles in online groups, grouping similar types of proteins in protein-interaction networks [8], and uncovering clusters that are separated by characteristics such as geographic location. Sometimes terms Cluster Analysis and community detection are misunderstood as different terms. Both terms are similar with a little difference. Community detection is a type of graph clustering where the data to be used for detection of clusters deals with relationships instead of features. Community structure is one of the important properties of complex networks. So we can say community detection is a clustering problem with specific attributes.

Table 1: Difference between Clustering and Community Detection

Clustering	Community Detection
Includes vectors	Includes vertices and links
Based on Similarity	Based on connectivity
It has metric structure	It has relational structure

## DIFFERENT DEFINATIONS OF COMMUNITY DETECTION ALGORITHMS

Community discovery problem is very similar to clustering problem of data mining. Community discovery is a clustering task of data mining which is done on the graphs. Till date many algorithms have been proposed by researchers which have different definitions

**Density Based Algorithms:** These algorithms are based on the topology of the network edges. According to density based algorithms, community is a group in which there many edges between vertices but there are fewer edges between groups. These algorithms divide the network into groups which have maximum number of edges in each group and minimum number of edges between the groups.

**Node Similarity Based Algorithms:** These algorithms define community as a group of nodes which are similar to each other but different from rest of the network. Similarity can be structural similarity, shortest path between nodes or location based similarity (topological information, nodal attributes define location similarity).

**Pattern Based Algorithms:** These algorithms try to identify the largest pattern (cliques) with large common nodes. These algorithms show better performance than density based algorithms as these don't rely only on numeric values.

**Link Centrality Based Algorithms:** Link centrality is based on two main features: number of nodes the link is

connecting and how likely these connections are to be used. Link between communities are very central and few. So they are likely to be used mostly. Girvan and Newman [26] defined edgebetweenness measure by considering the total number of shortest paths going through a link. Radicchi et al proposed edge centrality measure. It is defined as the ratio of the number of existing cycles containing the link of interest to the number of possible cycles given the existing links.

**Other Algorithms:** Many authors have used data compression technique [22] and considered community as a set of regularities in the network topology which can be used to represent the whole network in a better way. A community founded by these algorithms will have maximum compactness and minimum information loss.

## RECENT ALGORITHMS OF COMMUNITY DETECTION

- *Edge betweenness by Newman et al* [26] relies on the edge betweenness measure. It estimates the centrality of a link by considering the proportion of shortest paths going through it in the whole network
- Radicchi et al. proposed a variation called *Radetal*[38] based on link transitivity instead of edgebetweenness. This measure is defined as the number of triangles to which a given link belongs, divided by the number of triangles that might potentially include it.
- *Walktrap proposed by Pons and Latapy* [24] is based on the idea that a small random walk will stay inside the community from where it's originating because there are many links inside and few bridges leading outside
- *Fast Greedy developed by Newman et al* (Newman 2004) [11] is a greedy optimization algorithm for modularity. In it each node is initially in its own community and then, at every step, the algorithm groups two communities in order to maximize the gain of modularity.
- *Louvian (LV) proposed by Blondel et al* [33] is an optimization algorithm which gives an improvement over the Fast greedy algorithm.
- *Eigenvector Algorithm proposed by Newman* (Newman 2006) [34] is modularity-based, and it uses an optimization method inspired by graph partitioning techniques. It relies on the eigenvectors of a so-called modularity matrix, instead of the graph Laplacian traditionally used in graph partitioning
- *Label Propagation Algorithm. Proposed by Raghavan et al* (2007) [35] uses the concept of node neighborhood and the diffusion of information in the network to identify communities.
- *Markov Clustering (MCL) by Dongen* [35]. MCL is based on the idea that a random walk entering a dense cluster likely remains for a long time inside the cluster before switching between sparsely connected communities.
- *Commfind (CF) developed by Donetti and Muñoz* (2005) [36]. It combines the analysis of the Laplacianmatrix eigenvectors used in classic graph partitioning with a cluster analysis step. Instead of using the best eigenvector to iteratively perform bisections of the network, it takes advantage of the best ones. Communities are obtained by a cluster analysis of the projected nodes in this dimensional space.
- *Infomod (IND) was proposed by Rosvall and Bergstorm* (2007). It is based on a simplified representation of the network focusing on the community structure: a community matrix and a membership vector [22]. The former is an adjacency matrix defined at the level of the communities (instead of the nodes), and the latter associates each node to a community

- *Infomap (INP)* is another algorithm developed by Rosvall and Bergstorm (2008) [37]. The community structure is represented through a two-level nomenclature based on Huffman coding: one to distinguish communities in the network and the other to distinguish nodes in a community.
- *Expectation-maximization algorithm* by Newman and Leicht. In it bayesian inference is used to deduce the best fit of a given model to the data represented by the actual graph structure. The goodness of the fit is expressed by a likelihood that is maximized by means of the expectation maximization technique
- *Spectral algorithm* by Donetti and Mu~noz. In this method spectral properties of the graph are used. The idea is that eigenvector components corresponding to nodes in the same community should have similar values, if communities are well identified.

## APPLICATIONS OF COMMUNITY DETECTION

The study of detecting communities in complex networks has many practical applications. Its use has benefited several application fields such as sociology, communication, computer science, biology, physics etc.

- Detected communities are useful in the study of topology analysis, functional analysis and behavioural analysis of complex networks.
- Communities in biological networks can help in understanding basic mechanisms which control normal cellular processes and diseases pathologies.
- Clusters of customers with similar interests in the network can be used to make recommender systems for viral marketing to enhance the business [28].
- In adhoc networks nodes can be divided into communities which can help in generating compact routing tables.
- Community detection can help in easy visualization of complex graphs.
- Clusters of large graphs can be used to make large data structures to store huge graph data efficiently and to easily solve navigational queries related to that graph such as path search.
- Community discovery in World Wide Web can help in detecting link farms( a link farm is any group of [web sites](#) which [hyperlink](#) to every other site in the group)
- Identification of influential nodes of sub communities within large communities can help in predicting churns in telecommunication network.

## TOPOLOGICAL PROPERTIES OF COMMUNITY DETECTION ALGORITHMS

### Embeddedness

The embeddedness is one of the topological properties of community detection algorithms. It measures how much the direct neighbours of the node belong to its community. The embeddedness can be defined as the ratio of the  $k_{int}$  (internal degree) to the  $k$  (total degree) of the considerable node.

$$e = k_{int} / k$$

### Community Size

Community size is one of the most important characteristic of community structure. The community size can be identified by power law with the exponent  $\beta$  that ranges from 1 to 2. Community sizes are mostly heterogeneous, with several small communities and also there are some very large ones. For example, minimal community size in the real world networks is 2, but maximal community size varies based on the granularity of modelled system and class.

### Internal Transitivity

The internal transitivity is based on classic local transitivity and also averaged over nodes that locate inside the community. In general, local transitivity of the node depends on how the direct neighbours are interconnected. The internal transitivity can be defined as the number of links that are presents between direct neighbours divided by number of links that one may obtained if they were all interconnected.

The internal transitivity for any community  $C$  is formally defined as:

$$T(C) = \frac{1}{n_c} \sum_{i \in C} \frac{2 \times l(i)}{k_{inf}(i) [k_{inf}(i) - 1]}$$

Here,  $n_c$  represents the number of nodes in the community  $C$ .  $l(i)$  represents number of links among the neighbours of node  $i$  that belongs to same community.  $k_{inf}$  represents the internal degree of some node  $i$ . In the real world networks or artificial generated networks, distribution of internal transitivity may vary based on the community size in different ways.

### Scaled Density

The scaled density is also one of the topological properties of community detection algorithms. The density  $\rho$  of the community  $C$  can be defined as the ratio of links that actually contains  $m_c$  to number of links it contains if the nodes were connected. If it is an undirected network, then it is denoted as  $n_c(n_c - 1)/2$ . Here,  $n_c$  represents the number of nodes in community and so the result obtained will be

$$\rho = 2m_c / (n_c (n_c - 1))$$

The community density allows accessing cohesion of the community while comparing with the overall network access. The scaled density is the property that can be obtained through multiplying the density by community size.

$$\beta(C) = \rho(C) n_c = 2m_c / (n - 1)$$

If the community that is considered is a tree, then it has only  $m_c = n_c - 1$  links, and  $\beta(C) = 2$ . At the same time, considered community is clique (which is completely connected sub-network), then  $m_c = n_c(n_c - 1)/2$  and so  $\beta(C) = n_c$ . The scaled density allows characterizing the structure of any community. In general, some real world networks like communication networks and internet have tree-like communities.

### Average Distance

The average distance is the distance between two nodes represents to length of their shortest path. While averaging all the pairs of nodes in the community, it allows accessing the cohesion of that community. In the real world

networks, small communities ( $n_c \leq 10$ ) are supposedly small world over the whole network, then it is possible to increase the community average distance  $\bar{l}$  logarithmically with the help of community size. The average distance will increase in the larger communities but more slowly or even stabilizes for the classes of real world networks like internet or communication networks. The small average distance will be described by high density (social) and presence of hubs (Internet, communication), or both (information and biological).

### Hub Dominance

From the community structure perspective, the hub is one of the nodes that are connected to several other nodes that belong to the same community. In the community  $C$ , the presence of central hub can be assessed by using hub dominance measure and it corresponds to the following ratio:

$$h(C) = \max_i(k_{int}) / (n_c - 1)$$

The numerator is maximal internal degree that found in community  $C$ , where denominator is maximal degree theoretically possible which is provided by community size. Here, hub dominance reaches 1 when at least any one node is connected to all other nodes in community. The hub dominance reaches 0, if no nodes are connected. In the real world networks, behaviour of hub dominance depends on the class. While for the communication networks, hub dominance is close to maximum for all the community sizes which means hubs will present in all communities.

### FUTURE IN COMMUNITY DETECTION ALGORITHMS

Detecting clusters or communities in real world network is a problem of considerable practical interest. The community detection problem has plenty of challenges as it is highly related to the problem of clustering large heterogeneous datasets. Till date many researchers have proposed number of algorithms, but all the community detection algorithms are different from each other and are not clearly defined[3, 21]. So heterogeneity of different algorithms poses a challenge to community detection. Different networks (biological, social etc) have their own properties. This difference in properties as led to the unsolved question: which algorithm is suitable for which type of network?

Moreover these algorithms don't detect the same communities. So the problem is how to compare the performance of these algorithms. Actually the researchers are interested in following information.

- What type of information is used by the algorithm? A network can have different type of data: link attributes (weights, directions) node attributes, different types of links.
- What type of community produced ( partition, overlapped)
- The nature of communities the algorithm identifies.
- Community detection algorithms can be tested on following type of data:
- Real world networks.
- Artificial networks & Benchmarks

It is very difficult to test the different algorithms using real world data. It is very costly and time consuming to obtain real world data. Moreover complex networks have many properties such as average degree, shortest path, degree

distribution etc which are very difficult to be controlled in real world networks. Artificial networks provide solution to all these problems. They are widely used to compare the performance of different community detection algorithms. We can easily generate artificial networks with desired properties using generative models. But these cannot be substitute to real world data; instead can act as complement. The first benchmarks for testing these algorithms were developed by Girvan and Newman called as GN benchmarks. GN benchmarks are very simple to use. Many algorithms give good result with GN benchmarks as all communities identified by them are identical in size. GN benchmarks produce networks with poisson distribution but real world networks follow power law distribution. So GN benchmarks are not so fruitful in comparing community detection algorithms. Now days LRF benchmarks proposed by Lancichinetti et al have replaced the GN benchmarks [23]. These benchmarks can generate undirected and unweighted networks with mutually exclusive communities. More over there are many metrics to measure the quality of the communities detected by these algorithms. One popular metric is modularity [13]. Modularity measure has been used by many authors in their research to compare the various community detection algorithms. Few others are Rand Index (RI), Purity, Normalized mutual information (NMI) and F measure. Many researchers have even compared these measures to compare which one gives the best result for all community detection algorithms. Community detection in complex networks is a very challenging problem. Much work as been done in this field but still it is not clear which algorithm to be used in what situation.

## CONCLUSIONS

A complex network is a very young and promising interdisciplinary field whose roots lie in graph theory. The field of complex networks is helpful in understanding many complex phenomena's such as spam detection, protein interaction and spread of disease etc. Complex networks have many properties which have been studied by many authors in the past. Community detection is one of the fields of complex network which has gained a lot of attention in today's world. Many algorithms have been proposed by different researches but still many questions are unsolved.

## REFERENCES

1. S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75 { 174, 2010.
2. M. E. J Newman - The structure and function of complex networks, *SIAM Review* 45, 167-256 (2003)
3. Andrea Lancichinetti, Santo Fortunato, 2009, Community detection algorithms: a comparative analysis, ar Xiv: 0908.1062v1 physics soc-ph
4. Richardt J and Bornholdt S, 2006, Statistical mechanics of community detection, *Physical Review* E74, 016110
5. Jordi Duch and Alex Arenas, 2005, Community detection in complex networks using extremal optimization, *Physical review* E72, 027104
6. R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002)
7. D. Gibson, J. Kleinberg, and P. Raghavan, "Inferring web communities from link topology," in *Proceedings of HYPERTEXT*, 1998.
8. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi A-L. 2002. Hierarchical Organization of Modularity in Metabolic Networks. *Science* 297:1551–1555

9. Newman MEJ, Barabasi A-L, Watts DJ. 2006. The Structure and Dynamics of Networks: (Princeton Studies in Complexity). Princeton, NJ: Princeton University Press
10. Goh K-II, Cusick ME, Valle D, Childs B, Vidal M, et al. 2007. The human disease network. *Proc Natl Acad Sci USA*
11. Newman, M. E. J, 2004b. Fast algorithm for detecting community structure in networks. *Phys. Rev. E* 69, 066133.
12. S. White and P. Smyth, "A spectral clustering approach to finding communities in graphs," in *Proceedings of SDM*, 2005.
13. M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, 2004.
14. L. Danon, A. D'az-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, p. P09008, 2005.
15. P. Boldi and S. Vigna, "The webgraph framework i: compression techniques," in *WWW*, 2004
16. Pei Jie and Xiaofan wang. A review of community detection methods in complex networks [J] [Journal Article]. *Complex system and complex science* 2005. 2(3): p. 1-12.
17. Clauset, A., M. E. J. Newman, and C. Moore. Finding community structure in very large networks [Journal Article]. *Physical Review E*, 2004. 70(6): p. 066111.
18. Gulbahce, N. and S. Lehmann. The art of community detection [Journal Article]. *Bioessays*, 2008. 30(10): p. 934-938.
19. Pastor-Satorras, R. and A. Vespignani. Epidemic spreading in scale-free networks [Journal Article]. *Physical review letters*, 2001. 86(14): p. 3200-3203.
20. Montoya, J. M. and R.V. Sol?. Small world patterns in food webs [Journal Article]. *Journal of theoretical biology*, 2002. 214(3): p. 405-412.
21. Hubert L and Arabie P 1985 Comparing partitions *Journal of Classification* 2 193-218
22. Rosvall M and Bergstrom C T 2007 An information-theoretic framework for resolving community structure in complex networks *Proceedings of the National Academy of Sciences of the United States of America* 104 7327-31
23. Lancichinetti A, Fortunato S and Radicchi F 2008 Benchmark graphs for testing community detection algorithms *Phys Rev E* 78 046110
24. Pons P and Latapy M 2005 Computing communities in large networks using random walks *LNCS* 3733 284-93
25. Michelle Girvan, Mark Newman, "Community structure in social and biological networks", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821-7826, 2002
26. M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39 {54, January 2005.

27. Mark Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6), June 2004.
28. P. De Meo, A. Nocera, G. Terracina, and D. Ursino, Recommendation of similar users, resources and social networks in a social internetworking scenario, *Information Sciences*, vol. 181, no. 7, pp. 1285{1305, 2011.
29. R. Albert and A. Barab\_asi, Statistical mechanics of complex networks, *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47{97, 2002.
30. C. Gkantsidis, M. Mihail, and E. Zegura, "Spectral analysis of internet topologies," in *IEEE INFOCOM*, 2003.
31. S. Voulgaris, A.-M. Kermarrec, L. Massouli'e, and M. van Steen, "Exploiting semantic proximity in peer-to-peer content searching," in *IEEE FTDCS*, 2004
32. Z. Li and P. Mohapatra, "Impact of topology on overlay routing service," in *IEEE INFOCOM*, 2004.
33. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte "Fast unfolding of communities in large networks" *Journal of Statistical Mechanics: Theory and Experiment* (10), P10008, 2008.
34. M. E. J. Newman "Finding community structure in networks using the eigenvectors of matrices" Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109–1040
35. Van Dongen, S.: Graph Clustering Via a Discrete Uncoupling Process. *SIAM J Matrix Anal Appl* 30, 121-141 (2008)
36. Donetti, Luca and Munoz, Miguel. (2005) 'Improved spectral algorithm for the detection of network communities'. arXiv physics/0504059v1.